Coordinating Data Intensive Science Education and Training

September 9-11, 2015, at:

National Center for Ecological Analysis and Synthesis (NCEAS)

<u>Steering Committee</u>
Leah Wasser
Tracy Teal
Erin Robinson
Matthew B. Jones
Stephanie Hampton
Wendy Gram
Scott Collins

Matthew B. Jones
@metamattj
jones@nceas.ucsb.edu
0000-0003-0077-4738

What? Data-intensive* training

- **Goal: Collectively design a**
  - **coordinated strategy**
  - **for environmental science**
  - **to broadly improve skills necessary for data-intensive science**
  - **through education and training initiatives**

* Data management, software engineering, data science, statistics, modeling, informatics, …

Who?

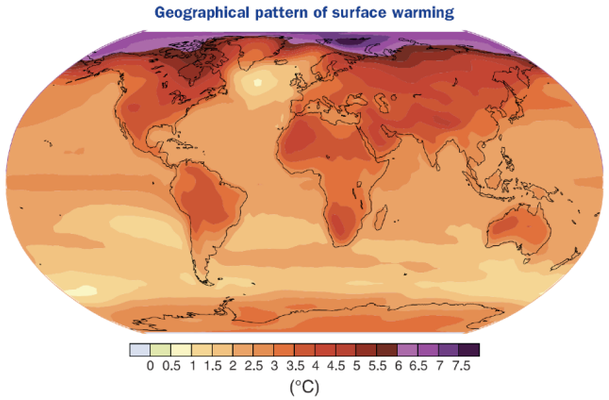
Workshop Contributors

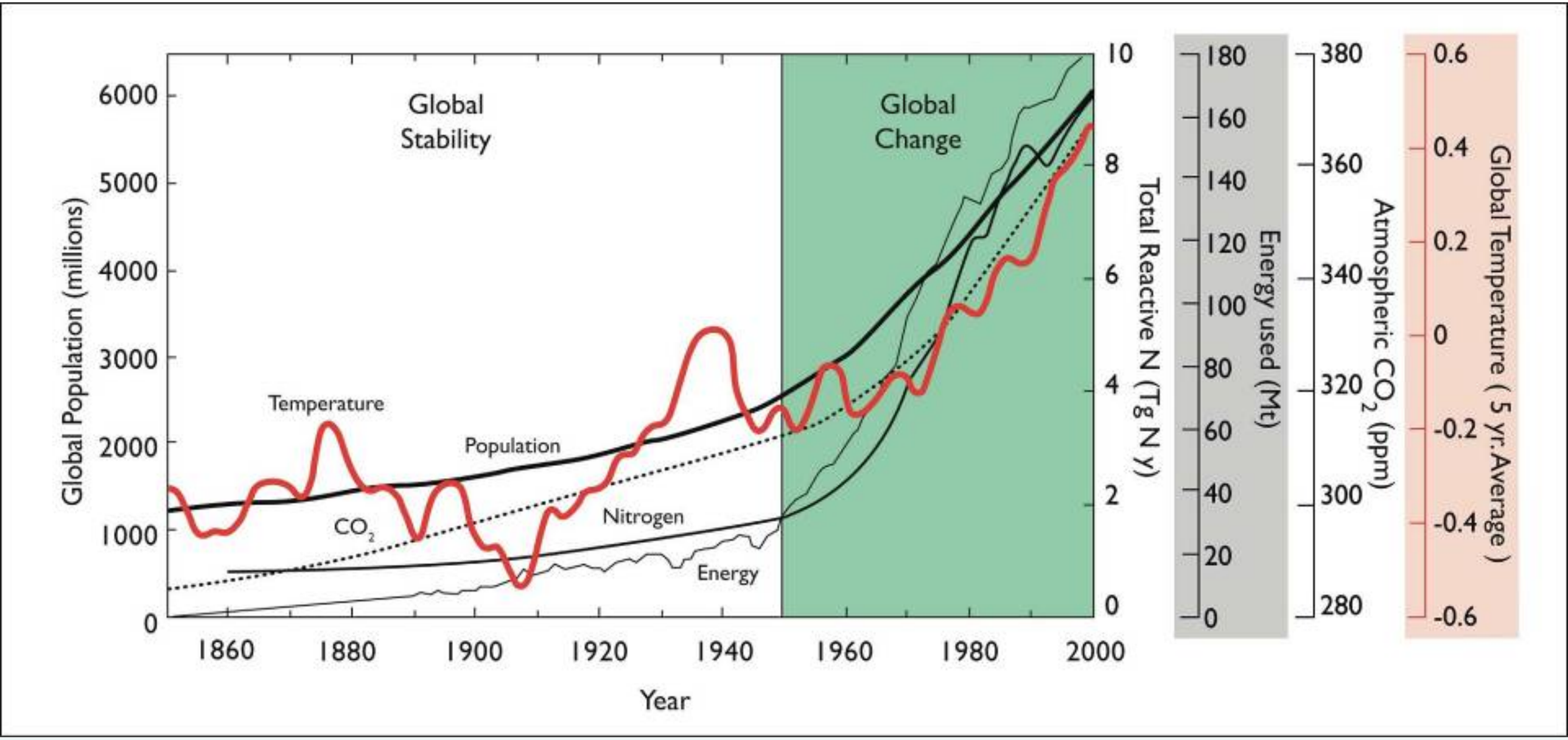
- Boettiger, Carl
- Bolker, Ben
- Bryan, Jennifer
- Budden, Amber
- Collins, Scott
- Fernandez, Denny
- Gram, Wendy
- Gross, Lou
- Hallett, Lauren
- Hampton, Stephanie
- Hernandez, Rebecca
- Jones, Matthew
- Robinson, Erin
- Schildhauer, Mark
- Shelley, Mary
- Smorul, Mike
- Supp, Sarah
- Teal, Tracy
- Wasser, Leah
- White, Ethan

Funding

National Science Foundation Grant # 1358900

Why?





Smith, Melinda D., Alan K. Knapp, and Scott L. Collins. "A framework for assessing ecosystem dynamics in response to chronic resource alterations induced by global change." Ecology 90.12 (2009): 3279-3289. doi:10.1890/08-1815.1

Why?

- Summarize existing workshop results and initiatives
- Examine differing approaches to training
  - e.g., science vs tools, …
- Generate vision for integrated training in data-intensive science

Agenda overview

- **Day 1**
  - Current training landscape
  - Curriculum catalog
  - Evaluation and assessment
- **Day 2**
  - Career pathways
  - Collaboration and sustainability
- **Day 3**
  - Manuscript writing

Collaboration tools

ecology

Collaborative notes

Etherpad:
http://notes.nceas.ucsb.edu/p/ds-workshop-2015

Version control

GitHub: https://github.com/NCEAS/ds-workshop-2015

Image credit: @ESAOpenSci

Products

- Training gaps and opportunities manuscript
- Prospectus for new collaborations

- Survey: the training lansdcape
- Curriculum catalog
- …

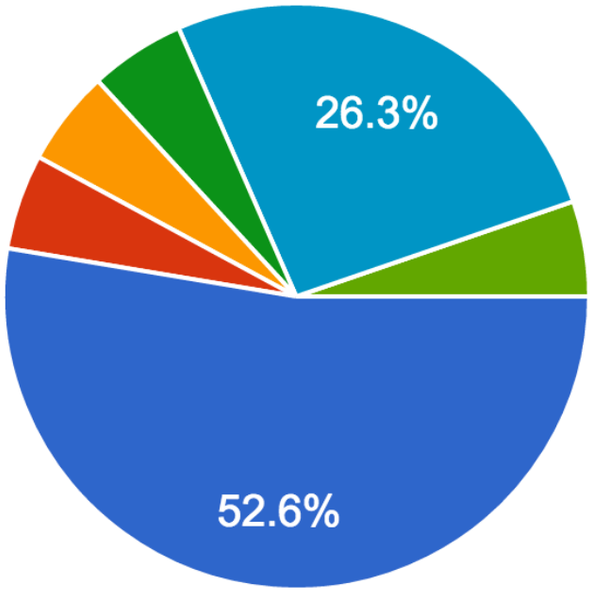
Training events survey

- 22 responses 8 Sep 2015
  - https://goo.gl/w5m7Na
  - https://github.com/NCEAS/ds-workshop-2015/tree/master/training-survey
- Rough information on the training landscape
- Starting point for more comprehensive survey
- Some activities mentioned more than once
- University classes under-sampled

Training Events

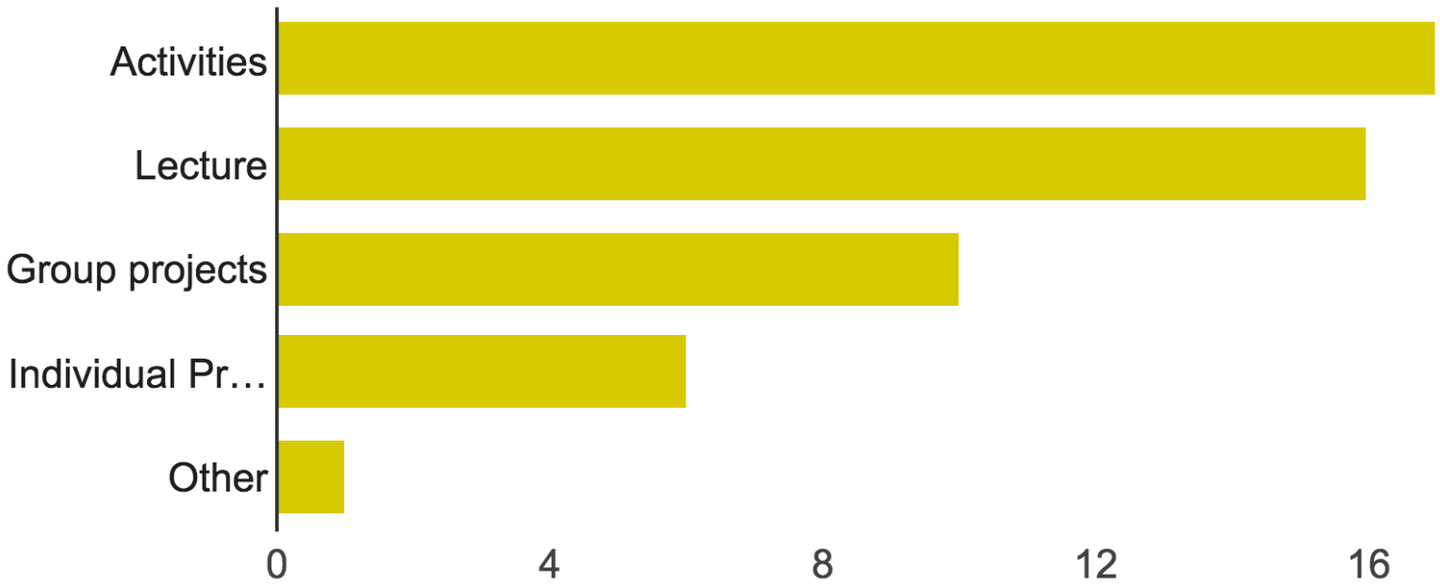
- Bayesian Workshop
- Computational Methods Applied to Biological Systems
- Computational Summer Institute
- Data Carpentry
- Data Carpentry Semester
- Data Science Curriculum
- Ecosystem Dynamics
- EEID meeting workshops
- ESIP Data Management Short Course
- Flux Course
- Managing Natural History Collections Data for Global Discoverability
- NEON Summer Institute
- NEON Data Skills
- NIMBioS/SAMSI/ESA Graduate Workshop on Current Issues in Statistical Ecology
- Open Science for Synthesis (OSS)
- Open Source Comes to Campus
- Programming for Biologists
- Reproducible Research Workshop
- Software Carpentry
- Software Carpentry Workshop
- Tutorials

Duration of Event



1-3 Days	**10**	52.6%
1 Week	**1**	5.3%
2 Weeks	**1**	5.3%
3 Weeks	**1**	5.3%
Quarter	**0**	0%
Semester	**5**	26.3%
Year	**0**	0%
Other	**1**	5.3%

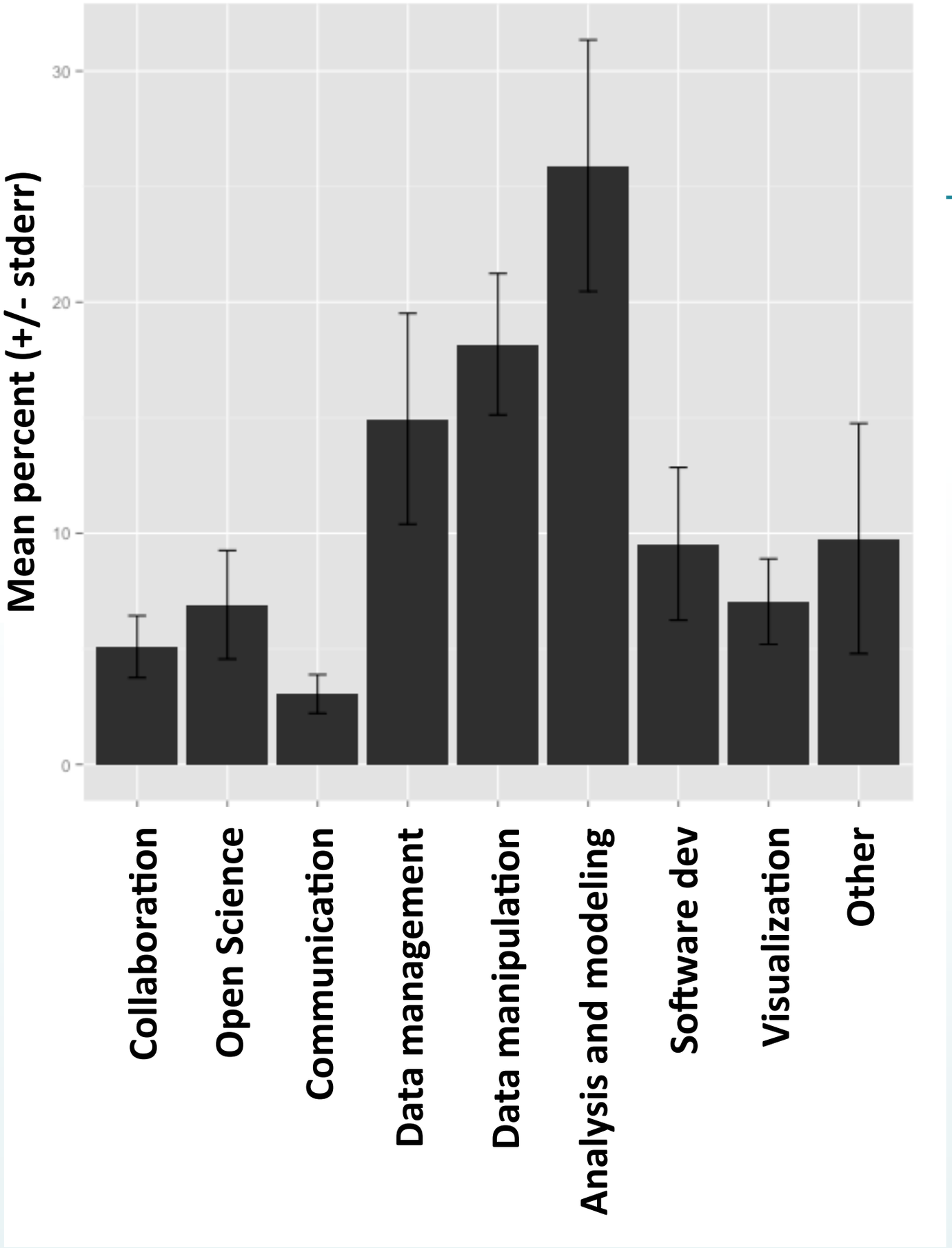
Event Format



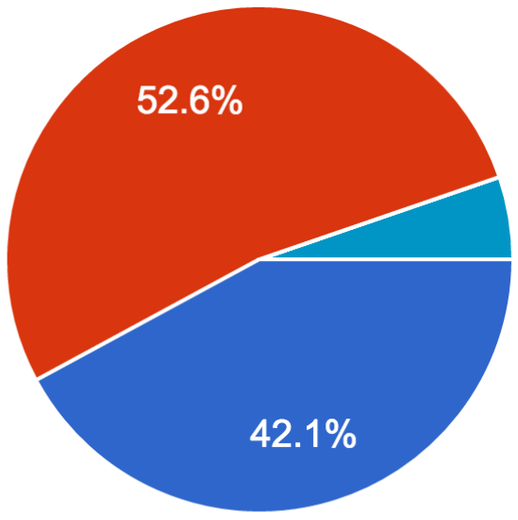
Activities	**17**	89.5%
Lecture	**16**	84.2%
Group projects	**10**	52.6%
Individual Projects	**6**	31.6%
Other	**1**	5.3%

Course Topics

- Other includes:
  - Reporting
  - Scientific Writing
  - Critical Thinking
  - Spatial data
  - HPC and R
  - HDF5
  - remote sensing
  - Sensor data
  - Synthesis



Organization Type

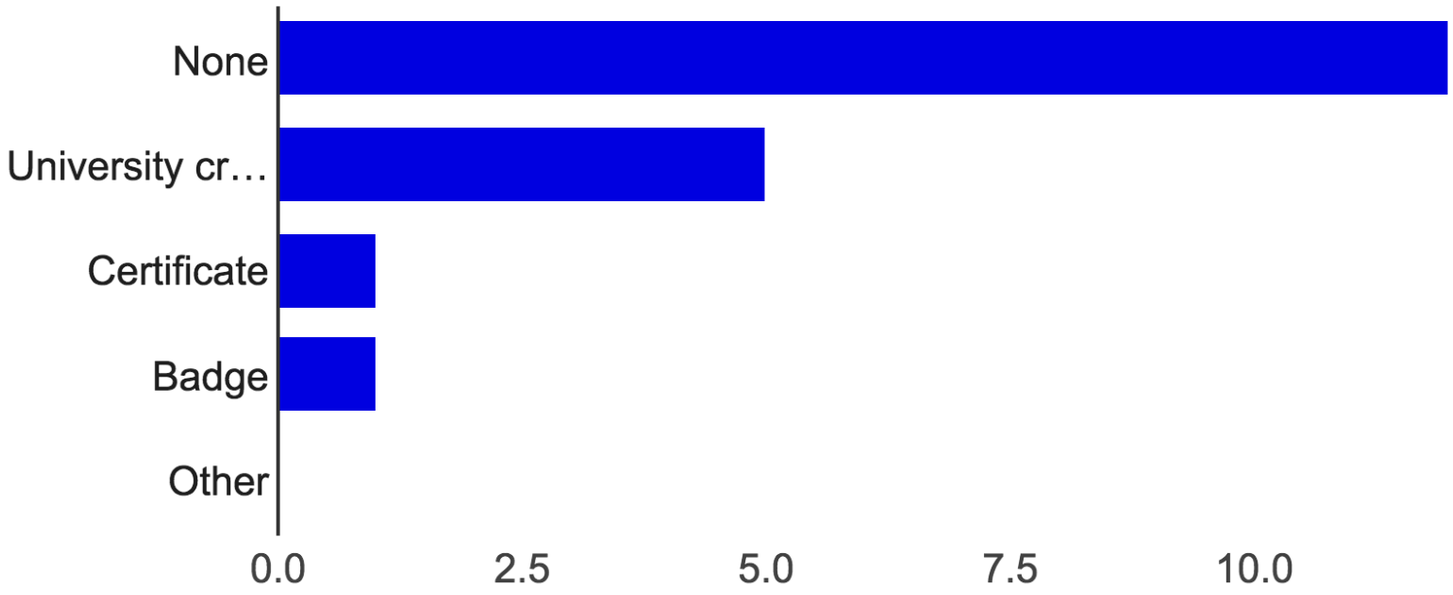


Not-for-profit	**8**	42.1%
University	**10**	52.6%
For-profit Corporation	**0**	0%
Federal agency	**0**	0%
State agency	**0**	0%
Other	**1**	5.3%

Selection Process

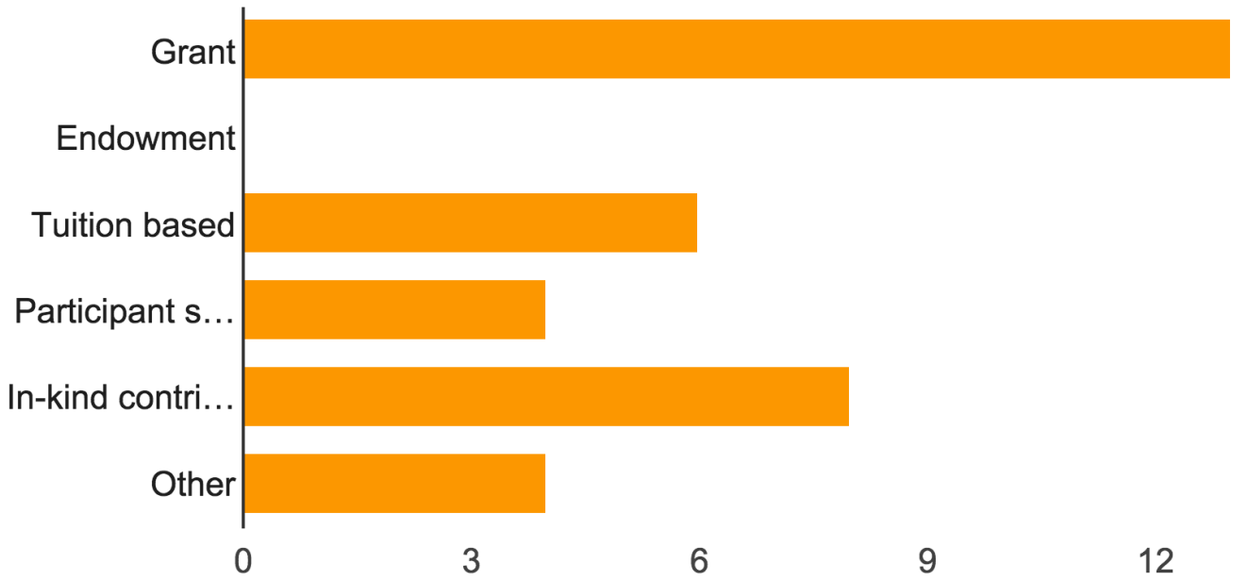
Application	**7**	36.8%
First come, first served	**11**	57.9%
Other	**3**	15.8%

Credit



None	**12**	63.2%
University credit	**5**	26.3%
Certificate	**1**	5.3%
Badge	**1**	5.3%
Other	**0**	0%

Funding Model



Grant	**13**	68.4%
Endowment	**0**	0%
Tuition based	**6**	31.6%
Participant self-funded	**4**	21.1%
In-kind contributions (e.g., teaching time)	**8**	42.1%
Other	**4**	21.1%

Brainstorming activity

- What are the critical gaps in data-intensive science training?
- What solutions can span these critical gaps?

- Write down 2-3 gaps, and 2-3 solutions
- Be concise!